SYNOPSIS

Enhancing the Performance of Aspect-based Sentiment Analysis and Opinion Summarization using Ensemble and Hybrid Models

Submitted by

SHINI GEORGE
(Reg. No: 17PHCSP009)

SUPERVISOR

Dr. V. Srividhya
Assistant Professor (SG)
Department of Computer Science
Avinashilingam Institute for Home Science and Higher Education for Women
Coimbatore - 641043

JANUARY 2023

1. INTRODUCTION

In today's digital environment, social media plays an important role in information exchange. Through reviews, it enables consumers to voice their opinions on any good, service, problem, or feature. With the expansion of Internet utilization, users contributed massive volumes of information in the form of blogs, reviews, and comments, making it difficult for users to identify meaningful information among the vast amount of data. As a result, techniques such as Topic Modeling, Sentiment Analysis and Opinion Summarization are useful in decision making for Individual Customers, Manufacturers, and Organizations. The Sentiment Analysis and Summarization is used as a technique to provide brief and relevant information about features related to the target domain.

Opinion mining (OM), also called sentiment analysis (SA), is a text mining technique whose goal is to analyze the thoughts, sentiments, and opinions about a subject or an entity. Many customers or users rely on review information to make decisions for movies, restaurants, shopping websites, and so on. Users may find it challenging to read and comprehend all the evaluations on the website. The goal of opinion mining is to investigate the public's attitudes, beliefs, and emotions concerning products, services, topics, and their attributes [1]. The Sentiment analysis procedure evaluates whether a user's view is positive, negative, or neutral. It is a study of feelings, thoughts, and points of view expressed through text [2]. The three levels of sentiment analysis are

- **Document Level**: Sentiment analysis at the document level divides a product or service's opinion into multiple sentiments. Based on document sentiment analysis, each document expresses opinions about a single entity.
- **Sentence Level**: For a product or service, a sentiment analysis at the sentence level determines if each sentence represents a positive, negative, or neutral opinion.
- **Aspect/Feature Level**: Identifying fine-grained opinion polarity towards a specific aspect of a given text is the goal. This type is used when there is a need for sentiments about desired aspect/feature in a review.

To manage the rapid expansion of digital document archives, techniques for automatically categorizing text documents based on its content are required. Unlabeled text data can be analyzed using topic modeling. Using text mining to identify hidden topics in data is a major revolution. Aspect Extraction is a process that could aid in finding the hidden aspects relating to the corpus. It will extract the major contents in a new article or will unearth the

pondered areas from a set of user reviews. Manual or automated methods can be used to identify aspects.

The great majority of publications concentrated on Conventional Summarization techniques. Summarization is the technique of summing up key contents from variety of information sources. This technique has become a vital part of everyday life. This method aims to produce a reduced representation of a text corpus that captures the essential meaning of the original material. The automatic text summarization is one the major challenges in Natural Language Processing. They are still facing with the task of producing a more Informative, Coherent, Fluent, and Concise Summary

Trip Advisor, Goibibo, Yatra, and other hotel booking websites use online reviews and customer feedback to analyze the facilities, quality, and service of restaurants, which is helpful for travelers. To provide concise and relevant information about features related to the target domain it has to extract relevant aspects-based opinions from large amounts of internet restaurant reviews. Furthermore, different summary methodologies necessitate distinct architectural approaches for constructing summaries. This work is intended to use Aspect-based sentiment analysis and text summarization for restaurant reviews in order to produce informative as well as concise summary.

2. REVIEW OF LITERATURE

Using 0.8 million Urdu tweets, Zoya et al., [3] analyze several topic modeling methods, LSA (Latent Semantic Analysis), PLSA (Probabilistic Latent Semantic Analysis), LDA (Latent Dirichlet Allocation), and NMF (Non-negative Matrix Factorization). To prevent the dominance of a particular topic in the corpus, tweets are retrieved using the Twitter API with a variety of hash tags. The tweet text was also pre-processed, three different datasets were created, and several features were retrieved to reflect documents on various n-grams. Moreover, using both qualitative and quantitative measurements, all of these strategies are compared and assessed on the dataset variations. In urdu twitter text, NMF outperformed strategies using TF-IDF feature vectors, but LDA outperformed techniques that merged short-text strategies into large pseudo documents.

In Annisa et al., [4] study, LDA is being used to retrieve topics from phrases that visitors frequently review in each topic. Using Latent Dirichlet Allocation, they seek to infer the subjects retrieved from hotel reviews. To increase the number of tourists who visit Mandalika, these issues are constantly discussed, which surely serves as feedback between hotel

management and guests. This model extracted eight topics using LDA based on the topic coherence.

Sazzed et al., [5] are primarily concerned with increasing the effectiveness of opinion categorization from unlabeled data. They provide SSentiA (Self-supervised Sentiment Analyzer), a self-supervised hybrid strategy for sentiment classification from unlabeled data that combines an Machine Learning(ML) classifier with a lexicon-based approach. LRSentiA (Lexical Rule-based Sentiment Analyzer), a lexicon-based method for forecasting the review's semantic orientation, as well as the prediction confidence score, is introduced first. By combining LRSentiA with an ML classifier, the hybrid technique SSentiA raises the macro F1 score for both binary and three-class sentiment analysis by 10% to 30%. The findings imply that SSentiA can greatly enhance sentiment classification performance in contexts where annotated data is missing. Furthermore, we show that SSentiA performs similarly to the completely labelled training dataset while employing 30%-60% annotated training data.

The detection of harmful remarks is recognized as a critical component for social media networks. To identify harmful remarks on social media networks, V. Rupapara et al., [6] employs an ensemble technique known as Regression Vector Voting Classifier (RVVC). The ensemble combines the support vector classifier and logistic regression under soft voting criteria. In order to assess the performance of the proposed approach, a number of experiments are conducted on balanced and unbalanced datasets. The synthetic minority oversampling approach is used to balance the uneven dataset. Additionally, two aspect extraction methods—Bag-of-Words(BoW) and term frequency-inverse document frequency are employed to assess the suitability of each BoW. Accuracy, precision, recall, and F1-score are used to assess the effectiveness of the proposed approach against a large number of machine learning classifiers.

Using five different models (Naïve Bayes, Linear Regression, Decision Tree, K-Nearest Neighbour, and SVM), AlGhamdi et al. [7] compared the effectiveness of three common ensemble methods for sentiment classification (Bagging, AdaBoost, and Stacking). Experiments were carried out on three various types of internet reviews, including restaurant, cellphone, and movie evaluations. The results showed that ensembles outperformed individual classifiers in general, with an average accuracy of 0.83 and recall of 0.82. The heterogeneous ensemble method known as stacking outperformed the homogeneous ensemble method, bagging, when the performance of the three ensemble techniques were examined. The results provide strong support to the idea that ensemble techniques, especially heterogeneous ones

using a robust classifier as a basis or Meta classifier, can improve sentiment classification enactment.

The categorization of unbalanced data sets is the subject of L. Wang et al., [8] study. Before reviewing the classification methods for unbalanced data sets from a variety of perspectives, including the data sampling approach, algorithm level, feature level, cost-sensitive function, and deep learning, a brief description of the type of data set is given. Additionally, synthetic minority oversampling, SVM, and KNN are discussed in the introduction. The merits and demerits of various tactics are then considered.

Numerous real-world datasets exhibit the class imbalance problem, which results in favoritism for the dominant class and subpar performance for the minority class. Desuky et al., [9] describe a unique hybrid method for dealing with such datasets. For under sampling, simulated annealing approach used, and for classification, SVM, decision tree, KNN, and discriminant analysis are used. Evaluate method on fifty-one real-world datasets and compare it to previous research as well. Proposed method outperforms previous strategies in terms of effectiveness and may thus be used to decrease misclassification in imbalance datasets. Their method was superior in seventeen datasets, whereas the Range-Controlled Synthetic Minority Over-sampling Technique (RCSMOTE) method was superior in just five datasets out of twenty eight. They also run Wilcoxon's signed ranked test, and the results show that the suggested method outperforms the other pre-processing strategies.

W. Liu et al., [10] study provides a two-stage summarization methodology, first extractive and then abstractive. Using the sentence similarity matrix (used for the first time) or the pseudo-title, which encompasses all the features of the relevant sentences, the relevant sentences (such as sentence position, paragraph position, etc.) are retrieved at the start of the process. To take the most important sentences from a document and extract coarse-grained sentences for sentence differentiation analysis. They use the beam search technique in the second step, which is abstractive, to reorganize and reorder the syntactic blocks of these retrieved sentences. In the following round, the newly created summary sentence will serve as a pseudo-summary. The final summary is a globally ideal pseudo-title. Extensive investigations on the matching data set have been conducted, and the findings demonstrate that our approach can get better outcomes.

Several ways to automated text summarizing has been presented to be able to abbreviate sections of the document. These algorithms produce decent results, but they still require

improvements in terms of sentence extraction reliability, redundancy, semantic links between sentences, and so on. A hybrid architecture combining a 2-layer Recurrent Neural Network (RNN) extractive model and a sequence-to-sequence attentional abstractive model is described by J. Pei et al., [11] in their research. The suggested approach combines the advantages of abstractive and extractive methods. Extractive models are used to extract pertinent information from a text, and abstractive models are used to summarize it. Based on preliminary experiments on real-world data, the proposed approach can produce competitive results for extractive, abstractive, and hybrid models.

3. PROBLEM STATEMENT AND OBJECTIVES

Identifying the aspects, classifying them based on opinions and creating a more informative and concise summary are performed through four phases namely Pre-Processing, Aspect Extraction, Sentiment Analysis and Summarization. Various algorithms have been proposed for each of these steps. However, existing ideas concentrate on improving just one of the phases at a time, and very few studies have examined each step individually before identifying ways to successfully combine the best approaches to create a system with increased performance. This research work attempts to solve the above-mentioned problem by proposing a system which enhances the performance of Aspect-based Sentiment Analysis and to create concise, clear, and coherent opinion summary using Ensemble and Hybrid Models.

To design a system that adheres to the above-mentioned problem, the primary objective of the research work is framed as, “To design and implement Aspect-based Summarization system for restaurant reviews by performing sentiment analysis using Ensemble machine learning techniques and Summarizing using Hybrid models”.

The specific objectives to meet the primary objectives were framed as follows.

- To determine the implicit aspects by using Topic Modelling methods.
- To design ensemble machine learning techniques for enhancing the performance of Sentiment Analysis.
- To create concise and coherent summary by using the hybrid models

4. METHODOLOGY

The research design for achieving the above-mentioned objectives is given in Fig. 1. There are four phases involved. The Restaurant Review dataset has been taken from TripAdvisor.com to implement this research work. The first phase is to prepare the data for

Pre-Processing. The second phase extracts the implicit aspect from the review dataset with the help of topic modeling techniques. To deal with imbalanced datasets, Synthetic Minority Oversampling Technique (SMOTE) is used. The third phase of this research work uses ensemble approaches for sentiment classification. The fourth phase presents two hybrid models that produce positive and negative summaries for each aspect.

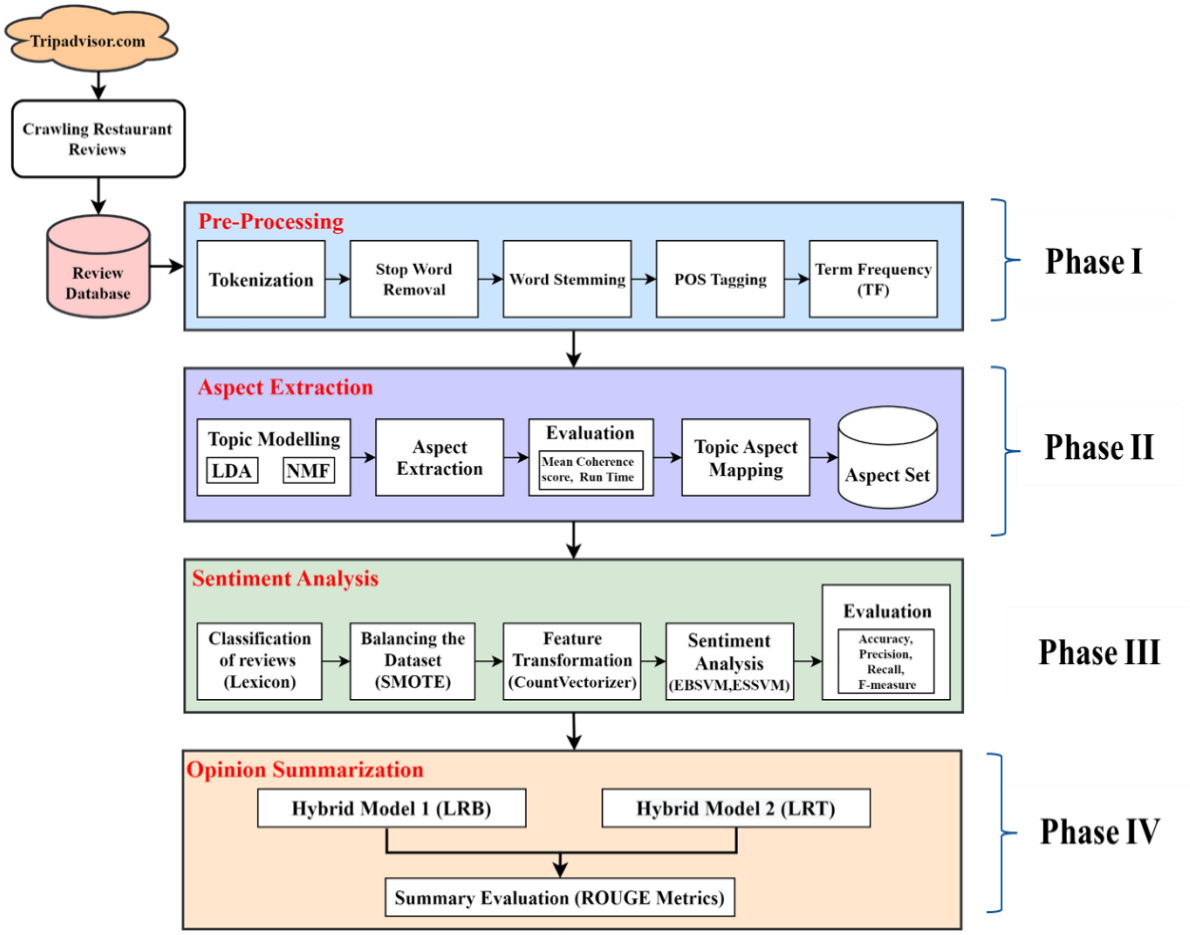

Fig. 1: Research Methodology

For the purpose of the proposed work, web crawlers were used to collect customer reviews for four restaurants in each of four metro regions from the TripAdvisor website. Totally 46,660 Reviews were taken from the period of January - December 2019.

4.1 Phase I - Preprocessing

A set of web-crawled review statements was fed into the data pre-processing stage. The purpose of this phase is to break down each document into individual words in order to represent each one as a feature vector. The preprocessing techniques such as tokenization, stop word removal, stemming, POS Tagging, and term frequency are used in this work.

i) Tokenization

It is the process of dividing a string sequence into components such as words, phrases, symbols, and other things called tokens. The words in this system are manually marked.

ii) Stop Word Removal

Stop words are common terms that have less significance than keywords. Stop-words, also known as language-specific functional terms, are common words that contain no information. Removal of stop words improves the understanding or accuracy of the Natural Language processing models.

iii) Stemming

Stemming, a normalization technique, reduces a list of tokenized words into shorter root words to avoid repetition. The process of stemming involves reverting inflected words to their stems or roots.

iv) Parts – of – Speech (POS)

The POS tagger is used to tag all review words. Each word in an online review statement is tagged using Stanford tagger. Every sentence in a customer review is tagged and saved in a text file.

v) Term Frequency (TF)

TF is a Count vectorization procedure that involves counting the number of occurrences of each word in a corpus. Term frequency refers to the number of times a specific word appears in a document compared to the total number of words.

$$TF = \frac{No.of\ Occurrences\ of\ a\ Word\ in\ the\ Document}{No.of\ Words\ in\ the\ Document} \quad (1)$$

4.2 Phase II - Aspect Extraction

Aspects are the important features rated by the reviewers. An Aspect is a group of terms that define a subtopic or theme in a specific domain. The purpose of this phase is to use topic modeling approaches to extract the Implicit Aspects from restaurant reviews. Topic Modeling is an unsupervised text mining approach for processing the large volume of data created in a corpus and extracting hidden and relevant aspects.

i) Topic Modeling

Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) are two Topic Modeling techniques that are used at this stage to extract characteristics from restaurant reviews. LDA is based on Probabilistic Graphical Models. It finds topics based on the words that are present in the document. Basic functionality involves taking N words and generating K topics, word distributions for each topic, and topic distributions for each document. NMF is a linear algebra algorithm used for identifying the latent structure that present in the data, which is represented as a Non-negative Matrix. For Restaurant Reviews, a fine-tuned aspect set was designed. Then, LDA and NMF are compared using performance indicators such as Coherence Score and Execution time for creating an Aspect Set based on the best topics.

ii) Topic Aspect Mapping

The pre-processed corpus is subjected to LDA and NMF in order to facilitate topic modeling and topic extraction. Due to the superior performance of NMF, its results were used in the subsequent stages. The algorithm then maps each topic to a particular aspect, and for each aspect, it checks the preprocessed corpus for matches with the topic words to extract aspect-specific language. The aspect-specific sentences will be written to.csv files for every aspect.

4.3 Phase III - Sentiment Analysis

The technique of sentiment analysis is used to extract opinions from reviews, and the results can be valuable. The methodology represents a framework of both single base and ensemble base classifiers. In this phase of research work, the following steps are performed for extracting the opinion from reviews.

i) Lexicon based Classification

The aspect-based review sentence is categorised as positive or negative using a lexicon-based technique. This work uses Vader package for Lexicon based classification. It is a sentiment analysis method based on rules. This stage determines the polarity of the review sentence and is stored in distinct .csv files for each aspect.

ii) Balancing the dataset with SMOTE

In this research, over- sampling approach is employed to balance the dataset. SMOTE is used to deal with imbalanced datasets. By adding synthetic samples based on feature-space

similarities between existing minority examples, the minority classes are oversampled. Synthetic data points can be generated by calculating the vector between one of the k neighbours and the current data point. The vector is then multiplied by a random number between 0 and 1. Adding this will create a new synthetic data point. According to the results, balancing the data minimises the likelihood of model overfitting, which occurs in an imbalanced dataset.

iii) Feature Transformation using Count Vectorizer

By calculating the frequency of each word in a file, the Count Vectorizer converts it into a vector. It transforms a set of text documents into a matrix of token counts. A document vector is generated from the file after feature selection. In this work, scikit-learn library in Python provides the Count Vectorizer.

iv) Single base Classifiers

To predict classification accuracy, base classifiers like Naive Bayes (NB), Decision Trees, and Support Vector Machines (SVM) are initially applied. When comparing the performance of the single base classifiers, SVM performs better than the other two.

v) Ensemble Classifiers

To boost the performance of Aspect-based sentiment categorization, two Ensemble models were developed.

a. EBSVM (Ensemble Bagging Support Vector Machine)

b. ESSVM (Ensemble Stacking Support Vector Machine)

a. EBSVM Classifier

In this methodology, a single ensemble classifier is created by combining the approaches of bootstrapping and aggregating machine learning. The multiple SVM is developed on each sub-sample of training data. Each SVM is built deep with sub-sampled training data to boost the classifier's performance, after that each SVM result is combined to get the optimum prediction. The accuracy of the model's prediction in the bagging strategy depends on how many SVM are utilised. Bagging with SVM is employed for homogenous Ensemble classification in this work. The number of SVMs utilised in the bagging strategy influences the model's prediction accuracy. With replacement, different sub-samples of sample data are chosen at random.

b. ESSVM Classifier

The work uses Stacking for heterogeneous categorization. An ensemble machine learning technique called stacking discovers the most effective way to combine the predictions from various successful machine learning models. It involves combining the predictions from multiple machine learning models on the same dataset. Stacking models consist of two or more base models, known as Level-0 models, and a meta-model that combines the predictions of the base models, called a Level-1 model. The top layer of the stack is more trained than the bottom layer as a result of feeding predictions from lower-level models to the subsequent model to construct the training dataset. The three single base ML classifiers employed in this model (Naive Bayes, Decision Tree, and SVM) serve as the layer one estimators. SVM also serves as the meta classifier.

4.4 Phase IV - Opinion Summarization

Aspect Based Text Summarization extracts the most important information from large document. There are two basic ways of summarising.

- Extractive Summarization creates a summary by extracting key sentences from the source document.
- Abstractive Summarization create a new text summary that doesn't include the source material. It is like humans build summaries.

There are two hybrid text summarization models created using the methodology. The extractive and abstractive summaries are included in the suggested models. For extractive summarising, the LexRank method is applied, and for abstractive summarization, Facebook's BART (Bidirectional and Auto Regressive Transformer) and Google's T5 (Text-to-Text Transfer Transformer) are both used. The two Hybrid models developed in this phase are LRB and LRT which is used for Opinion Summarization.

i) LRB Hybrid Model

The LRB hybrid model is built using LexRank and BART. The extracted summary produced by the LexRank algorithm is sent to BART (Bidirectional and Auto-Regressive Transformer). BART generates an abstractive summary from this extractive summary.

ii) LRT Hybrid Model

LexRank and T5 are used to construct the LRT hybrid model (Text-To-Text Transfer Transformer). T5 takes as input the extractive summary generated by LexRank. T5 generates an abstractive summary from the LexRank extractive summary.

5. EXPERIMENTAL RESULTS

A research work's success depends on evaluating the algorithms' performance. The algorithms proposed in this part of the research methodology are shown to meet the research objectives and to enhance the application process. In this research, Ensemble Machine learning algorithms and hybrid models were proposed to enhance the performance. The experiments were designed with the aim of studying the effectiveness of the algorithms and optimization methods proposed in each phase and were conducted in three stages.

Stage 1: Evaluate the topic modelling algorithms

Stage 2: Evaluate the single base classifier and ensemble classifiers

Stage 3: Evaluate the hybrid text summarization models

Stage 1: Despite the fact that topic models incorporate statistical analysis, each algorithm relies on different assumptions and has its own uniqueness. The interpretation of models still relies heavily on human judgment and domain expertise. The topics generated by the topic modeling algorithms in LDA and NMF are compared. As a result of the clear separation between all the identified themes in the NMF model, the findings obtained by NMF outperform LDA in general.

Stage 2: In this stage, the aspects which were derived from the previous stage are subjected to sentiment analysis. A comparison of the proposed Ensemble models against the base classifiers is carried out. The overall accuracy percentage of machine learning models on balanced databases for different aspects were calculated. The proposed ESSVM is more accurate than the EBSVM and base classifiers.

Stage 3: To create a concise, clear, and coherent summary for each positive and negative aspect, two hybrid approaches were proposed. Qualitative and quantitative analyses were used to evaluate two hybrid models. In addition to evaluating the quality of the models using the human-generated summary, the ROUGE score is used as a quantitative metric to measure the models' performance. It compares model-generated summaries to human-generated summaries based on content coverage. Scores were determined by comparing n-grams between system-

generated and reference summaries. The suggested summarization model was analyzed using ROUGE score. Efficiencies were measured by precision, recall, and F-scores in the ROUGE tool. The results of this stage, leads to the conclusion that the LRT Hybrid model outperforms LRB Hybrid model.

6. CONCLUSION

Increasing customer satisfaction requires analyzing the digital data to extract user opinions. A novel approach to aspect-based sentiment analysis using unstructured text was proposed in this work in order to generate relevant opinion-based features. Based on the dataset, LDA and NMF algorithms are applied to uncover hidden themes as "topics", which are then mapped to various aspects. From the above algorithms, NMF outperformed LDA and achieved higher coherence score for its topics. The model uses SMOTE to balance the dataset. Using machine learning techniques, an aspect-specific sentiment analysis was performed. To improve the performance, this work uses two Ensemble models EBSVM and ESSVM. According to the results, the suggested models excel all other individual classifiers as well as with the state-of-the-art approaches. On each aspect of the restaurant reviews, both ensemble models EBSVM and ESSVM shows highest accuracy of 96.1 % and 96.88 % respectively. This research also proposed two hybrid models (LRB hybrid model and LRT hybrid model) that produce concise, clear and coherent summaries for each positive and negative aspect (food, service, staff, ambience, and price). These two models were compared with human-generated summary, and their performance was rated using ROUGE scores. Resulting from the analysis, the LRT hybrid model is superior to the LRB hybrid model.

REFERENCES

[1] Golande, A., Kamble, R., Waghere, S., (2016), “An Overview of Feature Based Opinion Mining”, Intelligent Systems Technologies and Applications 2016, Advances in Intelligent Systems and Computing, vol 530. Springer, Cham. https://doi.org/10.1007/978-3-319-47952-1_51 .

[2] M. D. Devika, C. Sunitha and A. Ganesh, (2016),"Sentiment Analysis: A Comparative Study on Different Approaches," Procedia Computer Science, vol. 87, pp. 44–49, 2016.

[3] Zoya, S. Latif, F. Shafait and R. Latif, (2021) "Analyzing LDA and NMF Topic Models for Urdu Tweets via Automatic Labeling," in IEEE Access, vol. 9, pp. 127531-127547, 2021, doi: 10.1109/ACCESS.2021.3112620.

[4] Annisa, R., Surjandari, I. and Zulkarnain, (2019), “Opinion mining on Mandalika Hotel reviews using latent Dirichlet allocation,” Procedia Computer Science, 161, pp. 739–746, https://doi.org/10.1016/j.procs.2019.11.178.

[5] Sazzed, S. and Jayarathna, S. (2021), “SSENTIA: A self-supervised sentiment analyzer for classification from unlabeled data,” Machine Learning with Applications, 4, https://doi.org/10.1016/j.mlwa.2021.100026.

[6] V. Rupapara, F. Rustam, H. F. Shahzad, A. Mehmood, I. Ashraf and G. S. Choi, "Impact of SMOTE on Imbalanced Text Features for Toxic Comments Classification Using RVVC Model," IEEE Access, vol. 9, pp. 78621-78634, 2021, doi: 10.1109/ACCESS.2021.3083638.

[7] AlGhamdi, N. and Khatoon, S. (2021), “Improving sentiment prediction using heterogeneous and homogeneous ensemble methods: A comparative study,” Procedia Computer Science, 194, pp. 60–68. https://doi.org/10.1016/j.procs.2021.10.059.

[8] L. Wang, M. Han, X. Li, N. Zhang and H. Cheng, (2021), "Review of Classification Methods on Unbalanced Data Sets," IEEE Access, vol. 9, pp. 64606-64628, 2021, doi: 10.1109/ACCESS.2021.3074243.

[9] Desuky, A.S., Hussain, S. (2021), “An Improved Hybrid Approach for Handling Class Imbalance Problem”. Arab J Sci Eng 46, 3853–3864, https://doi.org/10.1007/s13369-021-05347-7

[10] W. Liu, Y. Gao, J. Li and Y. Yang, (2021), "A Combined Extractive With Abstractive Model for Summarization," in IEEE Access, vol. 9, pp. 43970-43980, 2021, doi: 10.1109/ACCESS.2021.3066484.

[11] J. Pei, R. Hantach, S. Ben Abbès and P. Calvez, (2020), "Towards Hybrid Model for Automatic Text Summarization," 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), 2020, pp. 987-993, doi: 10.1109/ICMLA51294.2020.00160.